

Theorie und Praxis von Hochschulrankings

Hornbostel, Stefan

Veröffentlichungsversion / Published Version

Sammelwerksbeitrag / collection article

Empfohlene Zitierung / Suggested Citation:

Hornbostel, S. (2007). Theorie und Praxis von Hochschulrankings. In *Amtliche Hochschulstatistik und Hochschulrankings : Beiträge zur wissenschaftlichen Tagung des Statistischen Bundesamtes am 9. und 10. November 2006 in Wiesbaden* (S. 6-13). Wiesbaden: Statistisches Bundesamt. <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-25282>

Nutzungsbedingungen:

Dieser Text wird unter einer Deposit-Lizenz (Keine Weiterverbreitung - keine Bearbeitung) zur Verfügung gestellt. Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Terms of use:

This document is made available under Deposit Licence (No Redistribution - no modifications). We grant a non-exclusive, non-transferable, individual and limited right to using this document. This document is solely intended for your personal, non-commercial use. All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.

Theorie und Praxis von Hochschulrankings

1 Einleitung

Von einer Theorie des Rankings zu sprechen ist fast ebenso schwer, wie über die Praxis zu berichten. Ersteres, weil es genau besehen gar keine Theorie des Rankings gibt, letzteres, weil nur wenige Rankings ihre Methoden und Verfahren so transparent machen, dass man ein Urteil über die Praxis fällen kann.

Die Grundidee eines Rankings ist – zumindest für Sozialwissenschaftler – eigentlich sehr befremdlich. Sie besteht nämlich darin, Informationen auf einem möglichst niedrigen Skalenniveau anzubieten. Anschaulich wird das, wenn man die zugrunde liegende Praxis im Sport betrachtet (vgl. Stölting 2003). Bei einem 100-m-Lauf etwa wird mit feinsten Messtechnik die benötigte Zeit auf Sekundenbruchteile genau ermittelt. Die Messung erfolgt auf höchstem Skalenniveau. Zur Siegerehrung werden dann aber drei der Läufer auf das Podest gebeten. Dort werden die Zeiten der drei Läufer auf einer ordinalen Skala dargestellt, man erfährt nur noch, dass A schneller als B und dieser schneller als C war. Unter alltäglichen Bedingungen wäre so etwas schlicht unsinnig. Wenn es etwa darum ginge den Bus noch mit einem Sprint zu erreichen, dann würden beispielsweise A und B als „gleich schnell“ gewertet werden (sie schaffen es noch), während C weit abgeschlagen nur noch die Rücklichter sehen würde. Sinnvoll erscheint das sportliche Ranking also nur, weil die Spielregeln i.d.R. vorsehen, dass es nur einen Sieger geben darf und ebenso auch nur eine Silber- bzw. Bronzemedaille. Diese Art der Komplexitätsreduktion ist ein sehr altes und weit verbreitetes Muster. Würdungen und Preise werden meist nur einer Person zuteil, die aus einer Gruppe ähnlich qualifizierter Kandidaten ausgewählt wird und Berufungslisten für Professoren sehen meist drei Rangplätze vor. Die Beispiele zeigen allerdings auch die Probleme: Gelegentlich wird der Nobelpreis geteilt, manche Berufungsliste sieht eine pari-passu Platzierung vor, kurz, manchmal scheint es unmöglich eine Rangfolge zu erzwingen. In den meisten Lebensbereichen verhält es sich sogar regelmäßig so, dass Rangplätze mehrfach besetzt sind. Kein Lehrer käme auf die Idee seinen Schülern Rangplätze zuzuordnen. Er konstruiert vielmehr ein Rating, in dem die Noten mehrfach vergeben werden.

Verglichen mit dem 100-m-Lauf muss ein Hochschulranking eine ungleich größere Komplexität abbilden. Das man dazu zum Ranking greift, lässt sich wohl nur historisch erklären, etwa derart, dass es nahe lag den Wettkampf der Football-Teams amerikanischer Universitäten einfach auf die akademischen Leistungen zu übertragen, um so ein wenig Übersicht in ein von extremen Qualitätsdifferenzen gekennzeichnetes Hochschulsystem zu bringen. Tatsächlich findet sich auch kaum noch ein reines Ranking. Meist sind den Rangplätzen zusätzliche metrische Informationen beigelegt, die es erlauben Distanzen abzuschätzen und/oder es werden Ranggruppen gebildet, so dass aus dem Ranking ein Rating wird. Gleichwohl ist damit ein Grundproblem noch nicht gelöst. Wir sind es gewohnt, Unterschiede daraufhin zu betrachten, ob sie substantiell – statistisch gesprochen signifi-

*) Prof. Dr. Stefan Hornbostel, iFQ – Institut für Forschungsinformation und Qualitätssicherung, Bonn.

kant – sind. Statistisch sind dabei allerdings schnell Grenzen erreicht. Man kann zwar in Anlehnung an die Stochastik bei geeigneten Daten (etwa studentischen Urteilen) Konfidenzintervalle konstruieren, deren Überschneidungsfreiheit eine substantielle Differenz signalisiert, streng genommen haben wir es aber meist mit Vollerhebungen und nicht mit Stichproben zu tun. Ignoriert man den wahrscheinlichkeitstheoretischen Hintergrund, lassen sich Vertrauensintervalle aber immerhin als eine Art „Sicherheitspuffer“ behandeln, der vor vorschneller Differenzbildung schützt. Aber schon ein so einfacher Schritt zeigt, wie schnell die Komplexität der Darstellung anwächst, denn ein solcher Sicherheitspuffer gilt individuell für die Untersuchungseinheiten, erzwingt also eine ganz andere Darstellung als eine Rangtabelle und verunmöglicht eine klare Siegerehrung. Meist behilft man sich daher mit Ranggruppen oder schlicht mit Quartilbildung, was zur Folge hat, dass an den Klassengrenzen ähnlich gute Einheiten mal in die Spitzen-, mal in die Mittelgruppe fallen. An die Grenze zur Irreführung geht es allerdings, wenn man dem Nutzer von Rankings die originale Information über Distanzen vorenthält und ihm stattdessen eine Prozentskala anbietet, auf der die beste Einheit mit 100 % verzeichnet ist und alle anderen daran gemessen werden. Optisch werden so auch kleinste Unterschiede zu beeindruckenden Prozentwerten aufgeblasen.

Die Frage nach der Erheblichkeit von Unterschieden wird jedoch in allen Rankings eher intuitiv, als methodisch begründet entschieden.

2 Warum und für wen braucht man Rankings?

Rankings sind im Trend. Noch vor 16 Jahren verzeichnete das Educational Rankings Annual 1 500 solcher Studien, inzwischen sind es über 4 600 und das ist vermutlich eine Unterschätzung der weltweit existierenden Hochschulrankings (vgl. Nickl 2006, S. 7). Inzwischen findet auch ein intensiver internationaler Austausch über nationale und weltweite Rankings statt (vgl. insbes. IHEP 2007; Merisotis/Sadlak 2005; in Higher Education in Europe No. 4, 2002 findet sich ein Überblick zu international vergleichenden Rankings, u. a. Federkeil, Jobbins, Ledesma, Merisotis; vgl. auch Federkeil 2006; Buela-Casal u. a. 2007).

Rankings sollen entweder Informationen für Akteure bereitstellen, die Auswahlen zwischen unterschiedlichen Angeboten treffen müssen oder sie werden mittel oder unmittelbar der Ressourcenallokation genutzt. Sie dienen also letztlich einer Stimulation des Wettbewerbs, verbunden mit der Hoffnung, dass so eine beständige Qualitätsverbesserung erreicht werden kann. Entsprechend waren Rankings im US-amerikanischen Hochschulsystem mit seinem hohen Anteil von privaten Anbietern und seiner starken Marktorientierung sehr früh verbreitet.

Unmittelbare Mittelzuweisungen in Verbindung mit Rankings führen meist zu heftigen Anpassungsreaktionen, wobei nicht unbedingt nur die intendierten Effekte einer Qualitätsverbesserung eintreten, sondern auch eine intelligente Optimierung von Indikatoren betrieben wird. Besonders problematisch wird dies, wenn „nicht-anreizkompatible Indikatoren“ benutzt werden. So wird z. B. in Deutschland regelmäßig die absolute oder relative Zahl der Promovierenden als Indikator präsentiert. Ohne eine Aussage zur Qualität dieser Promotionen oder wenigstens des beruflichen Verbleibs der Absolventen ist

dies eine Aufforderung zur Senkung der Qualitätsstandards, denn wenn die schiere Menge „belohnt“ wird, ist der schnellste, einfachste und effizienteste Weg zum Ziel eine Senkung der Anforderungen.

3 Worüber informieren Rankings?

Selbstverständlich über die Qualität von Lehre und Forschung, so könnte man meinen. Tatsächlich ist das keinesfalls die Regel. Je nach Ranking wird über Reputationsurteile, Befragungsergebnisse von Studierenden, die Studiendauer, Noten, über Forschungspreise, bibliometrische Befunde, Patentierungsaktivitäten, Drittmiteleinwerbungen, amtliche Daten oder auch anderweitig erstellte Evaluationsergebnisse berichtet und diese Informationen entweder zu einem Index verrechnet oder getrennt präsentiert. Selten wird diesen Angaben auch in verlässlicher Weise eine Inputgröße zugeordnet, so dass sich tatsächlich Effizienz oder Belastung beurteilen lässt. Reine Quantitäten sind zwar eine Information über die Sichtbarkeit einer Einrichtung in der akademischen Welt, sie verraten aber nichts über die Leistungen, die hinter diesen Mengenangaben stehen. Das gilt für Drittmitlebenso wie für Nobelpreise.

Rankings müssen, wenn sie ihrem Anspruch auf Lieferung belastbarer Qualitätsinformationen gerecht werden wollen, eine Reihe von Problemen lösen:

- Sie müssen geeignete Untersuchungseinheiten bestimmen, auf die sich die Informationen beziehen.
- Sie müssen das sehr vage Konstrukt „Qualität“ durch angemessene Indikatoren operationalisieren.
- Sie benötigen für eine solche Operationalisierung eine geeignete Datenbasis.
- Und sie sollen schließlich zielgruppenspezifisch aufgebaut sein.

Das erste Problem, geeignete Untersuchungseinheiten zu definieren, wird sehr unterschiedlich gehandhabt. Benutzt man die ganze Hochschule als Einheit, dann lässt sich keine sinnvolle Aussage über die Leistungen in einzelnen Fachgebieten treffen: Eine gute Physik bedeutet nicht automatisch, dass auch die Germanistik am Ort exzellent ist. Schlimmer noch: Gerade Forschungsindikatoren sind in der Regel fachspezifische Indikatoren. Die Drittmiteleinwerbungen differieren nach Fachgebieten erheblich, die Publikationsweisen und die zugehörigen Zitationsmuster ebenso (vgl. CHE ForschungsRanking, DFG Förder-Ranking). Benutzt man eine ganze Hochschule als Untersuchungseinheit, entstehen notwendigerweise Kompositionseffekte. Für eine Hochschule mit einem Schwerpunkt in den Lebenswissenschaften werden sich immer deutlich mehr Artikel in Fachzeitschriften und auch deutlich höhere Zitationen, mehr Nobelpreise und mehr „highly cited papers“ nachweisen lassen, als für eine Hochschule mit Schwerpunkt in den Geistes- und Sozialwissenschaften. Einerseits, weil in den Geisteswissenschaften auch Bücher und Sammelbände verfasst werden, andererseits, weil diese Literatur in den internationalen Datenbanken (wie Web of Knowledge oder Scopus), die üblicherweise genutzt werden, kaum erfasst ist. Um eine ganze Hochschule zu beurteilen, muss man daher von den fachlichen Einheiten ausgehen und die Ergebnisse aggregieren.

Allerdings sind solche Einheiten nicht leicht zu definieren. Geht man von organisatorischen Einheiten (Instituten, Fachbereichen etc.) aus, hat man es mit historisch gewachsenen Institutionen zu tun, deren Zuschnitt von Universität zu Universität variiert. So kann die Didaktik mal in einem eigenen Institut untergebracht sein, mal Bestandteil der jeweiligen Fachdisziplinen sein, die Rechtsmedizin mal den Rechtswissenschaftlern, mal den Medizinerinnen zugeordnet sein, ein Stadtsoziologe mal zu einem sozialwissenschaftlichen Institut, mal zur Raum- und Regionalplanung gehören. Von derartigen Organisationseinheiten auszugehen, hat sicherlich den Vorteil, dass sich der zugehörige Input recht genau ermitteln lässt und dass handlungsfähige Einheiten Adressat eines Ranking sind. Der Nachteil besteht u. U. in der fachlichen Heterogenität und damit eingeschränkter Vergleichbarkeit. Zu anderen Ergebnissen kommt man daher, wenn man artifizielle Einheiten bildet, wie sie sich z. B. ergeben, wenn man die Fachgebietsklassifikationen von Literaturdatenbanken oder die Klassifikationen von Drittmittelgebern übernimmt. Im ersten Fall werden Publikationen, je nachdem in welcher Zeitschrift sie erschienen sind, Fachgebieten zugeordnet, im zweiten Fall sind bspw. die Fachkollegien der Deutschen Forschungsgemeinschaft (DFG), in denen ein Antrag verhandelt wird, für die fachliche Zuordnung maßgeblich. Schließlich wird zur Zeit vom Wissenschaftsrat ein Verfahren – analog zum britischen Research Assessment Exercise (RAE) – praktiziert, in dem die Hochschulen selbst homogene Untersuchungseinheiten definieren und das zugehörige Personal benennen. Zu echten Fehlern kommt es, wenn die unterschiedlichen Datenquellen beliebig gemischt werden, denn sie beziehen sich jeweils auf unterschiedliche Personenkreise. Selbst bei einer trennscharfen Bestimmung der Berichtseinheit bleibt zu klären, über wen berichtet werden soll: Die Leistungen der Professoren?, des gesamten wissenschaftlichen Personals, Drittmittelbeschäftigte eingeschlossen?, oder nur des forschungsaktiven Personals, wie es im britischen RAE vereinbart ist? Insbesondere bei internationalen Vergleichen wirft die Bestimmung einer geeigneten Personenbasis erhebliche Probleme auf, je nach dem, wie stark zwischen Lehr- und Forschungsaufgaben differenziert wird.

Das zweite Problem – die Operationalisierung von Lehr- und Forschungsqualität, ist sicherlich das komplizierteste (vgl. Neidhardt 2006). Lehr- oder Forschungsqualität ist eine soziale Konstruktion, keine einfach vorfindliche Entität. Entsprechend werden auf indirektem Wege schon einmal gefällte Urteile von kompetenten Peers als Anhaltspunkte für Qualität benutzt. Drittmittel, die nach fachlicher Begutachtung vergeben werden, Publikationen in Zeitschriften mit Review-Verfahren, Zitate, Preise oder Forschungsaufenthalte enthalten solche „geronnenen“ Peer-Urteile. Befragungen von Studierenden oder Professoren ermitteln hingegen aktuelle, subjektive Beurteilungen oder Zufriedenheiten (vgl. Hornbostel 2001; Kromrey 2001). Sie sind insofern schwerer zu interpretieren, als die zugrunde liegenden Aspirationsniveaus nicht immer klar oder vergleichbar sind. Ein ganzer Kranz von Daten beschreibt Voraussetzungen oder Potentialitäten von Lehr- und Forschungsprozessen: Studienzeiten, Ausstattungen, Bibliotheksmittel, Betreuungsrelationen, Nutzflächen etc. geben einen nutzbaren Rahmen an. Wie sich dieser Rahmen aber in der universitären Lebenswelt eines Studierenden darstellt, ist höchst variabel. Betreuungsrelationen markieren bspw. bestenfalls Problemzonen, sie sagen aber kaum etwas über die tatsächlich realisierte Betreuung in den verschiedenen Phasen eines Studiums aus. Reputationserhebungen sind im Grunde gar keine Qualitäts-

messung, sondern einfach eine „Bekanntheitserhebung“. Wichtig sind solche Informationen nur, weil sie in Gestalt einer „self-fulfilling prophecy“ Differenzen stabilisieren und jenseits tatsächlicher Leistungen dazu führen können, dass z. B. Absolventen von Eliteeinrichtungen – die möglicherweise wenig überzeugende Lehrleistungen erbringen – dennoch einen realen Vorteil bei der Stellensuche haben, weil alle unterstellen, dass eine Eliteeinrichtung auch eine hervorragende Ausbildung liefert (vgl. Kirp 2003; Newman u. a. 2004). Manche Rankings (z. B. Times Higher Education World University Rankings [THES] 2006) stützen ihren Gesamtscore zu 40 % auf derartige Reputationserhebungen. Solche Informationen sind durchaus hilfreich, wenn man eine ordentlich gezogene Stichprobe aus einer klar definierten Grundgesamtheit hat (etwa alle deutschen Professoren), in der Praxis handelt es sich aber meist um ein (preiswertes) Spektakel, ohne Aussagekraft. Im World University Ranking erfährt man zu den ausgewählten Befragten nur, dass weltweit gut 4 000 Wissenschaftler und Personen aus der Wirtschaft und dem öffentlichen Sektor befragt wurden. Derartige Befragungen haben bestenfalls den Wert von Internet Polls, sie lassen mit der (nicht dokumentierten Auswahl) Raum für jede Art von Manipulation.

Es würde den Rahmen sprengen, die Konstruktions- und Messprobleme der einzelnen Indikatoren im Detail zu erörtern (vgl. Hornbostel 1997). Wichtig ist aber sich zu vergegenwärtigen, dass jeder dieser Indikatoren eine eigene Mess- und Fehlerlehre benötigt. Das schmälert nicht den Nutzen solcher Indikatoren, setzt aber ein erhebliches Interpretationswissen bei den Nutzern voraus.

Anschaulich wird das anhand des dritten Problems – einer geeigneten Datenbasis. Pars pro Toto seien hier die angesichts immer leichter Verfügbarkeit häufig genutzten Publikations- und Zitationsindikatoren genannt. Die Datenbasis ist meist der Science Citation Index (SCI). Je nachdem, ob man Personen sucht (die zu einem Institut gehören) oder einfach nach institutionellen Adressen fragt, erhält man unterschiedliche Ergebnisse. Der Grund liegt in Fehlern und vor allen Dingen uneinheitlichen Schreibweisen von Autoren und Institutsadressen. Im internationalen Vergleich kann das dramatische Verzerrungen zur Folge haben (vgl. van Raan 2005). Eine Nutzung dieser Informationsquelle setzt daher regelmäßig umfangreiche Bereinigungsarbeiten voraus (die selten dokumentiert werden). Ebenso beeinflusst die Indikatorenkonstruktion das Ergebnis: Wird jeder Artikel voll gezählt oder werden Autorenbruchteile – entsprechend der Zahl der Koautoren – zugerechnet (fraktional counting) (vgl. Gauffriau/Larsen 2005)? Werden nur Erst- oder Letztautoren berücksichtigt oder alle Autoren? Wird ein festes oder variables Zitationsfenster gewählt? Werden nur Source-Items, d. h. Beiträge aus Zeitschriften, die im SCI ausgewertet werden oder auch Non-Source-Items, d. h. Beiträge, die zitiert werden, aber nicht in einer SCI Zeitschrift erschienen sind, berücksichtigt? Erfolgt die Zuordnung der Publikationen aufgrund der letzten institutionellen Zugehörigkeit des Autors oder aufgrund der Korrespondenzadresse, die der Autor angegeben hat? Werden einfach Zitate gezählt oder werden die Zitationen anhand der Durchschnittswerte im Fachgebiet normalisiert (vgl. van Raan 2004, S. 31)? Keine dieser Möglichkeiten ist einfach richtig oder falsch, es sind vielmehr unterschiedliche Modellierungen von Impact. Auch eine gemeinsame Datengrundlage sichert keineswegs gleiche Ergebnisse. Laien sind dann häufig erstaunt darüber, dass trotz gleicher Datenbasis unterschiedliche Ergebnisse entstehen (vgl. z. B. Lebherz u. a. 2005, S. 206). Zu Recht weist daher Weingart (2005)

darauf hin, dass ein unreflektierter Umgang mit bibliometrischen Indikatoren zu ausgesprochen inadäquaten Schlussfolgerungen führen kann.

Zielgruppenspezifität – das letzte Problem – ist ein Zauberwort, das im Zusammenhang mit Rankings immer wieder fällt. Zu Recht wird unterstellt, dass unterschiedliche Adressaten (Studienanfänger, Promovierende, Hochschulverwaltungen, Ministerien usw.) sehr unterschiedliche Informationsbedarfe haben. Das spricht für eine Vielzahl sorgfältig ermittelter Indikatoren, große Freiheit in der Auswahl und Kombination von Indikatoren, elaborierte Verfahren der Indexbildung und ausführliche Interpretationshilfen. Auf der anderen Seite besteht der Sinn von Rankings aber gerade in der Komplexitätsreduktion, in der Präsentation weitgehend voraussetzungslos nutzbarer, einfacher Information. Dieses Paradoxon führt dazu, dass zwei Varianten von Rankings auf dem Markt gehandelt werden: Die erste Variante setzt auf Einfachheit. Dazu werden meist verschiedene Indikatoren zu einem Score verrechnet. Algorithmen und Gewichtungen werden mehr oder weniger willkürlich durch den Rankinganbieter festgelegt, der entscheidet, was für die Zielgruppe geeignete Informationen sind. Diese Entmündigung des Konsumenten, euphemistisch als Delegation von Teilen des Entscheidungsprozesses umschrieben, wird theoretisch mit sehr viel Modellplatonismus gerechtfertigt (vgl. beispielhaft Bayer 2004) und unterschlägt die erheblichen Interpretationsprobleme, die hinter der Pseudoobjektivität eines „Gesamt-Scores“ unsichtbar wird. Der andere Weg – z. B. das CHE Ranking – verzichtet weitgehend auf die Verrechnung von Indikatoren und bietet die Kombination von Indikatoren zu einem individuellen Ranking an (und auf Nachfrage auch Sonderauswertungen, die Detailanalysen ermöglichen). Dieses sachlich vorbildhafte Vorgehen erzeugt aber erstens keine Siegerliste und stellt zweitens erhebliche Anforderungen an den Nutzer, der sich für eine Auswahl von Indikatoren mit deren Aussagekraft auseinandersetzen muss.

4 Resümee

Rankings – auch methodisch anspruchslos – können in einer qualitativ stark differenzierten Hochschullandschaft zwischen oben und unten unterscheiden und sei es nur in Gestalt sich selbst stabilisierender Reputationsurteile. Ungleich schwieriger wird es, wenn die Differenzen zwischen den Bewertungseinheiten weniger groß sind. Die Anforderungen an die methodische Solidität eines Rankings und damit auch die Kosten steigen enorm. Diese Situation haben wir in Deutschland. Ausgehend von einem formal gleichrangigem Institutionenensemble haben sich in den letzten Jahrzehnten immer stärker Stichworte wie Wettbewerb, Exzellenz, Profilbildung und öffentlich verhandelte Leistungsunterschiede in den Vordergrund geschoben. Dieser Trend wird sich auch kaum ändern, zumal Lehrpreise, leistungsorientierte Mittelvergabe, W-Besoldung, Studiengebühren, gezielte Studienortswahl, aber auch internationale Konkurrenz um Köpfe, den Bedarf an soliden Informationen über Leistungen in Forschung und Lehre weiter erhöhen (vgl. Lücke 2006; vgl. Hornbostel 2006). Dieser Informationsbedarf lässt sich mit einfach gestrickten Rankings nicht befriedigen. Der in internationalen Rankings nach wie vor vorherrschende Trend, ganze Hochschulen mit einem Index, der aus unterschiedlich gewichteten, einfach zu erhebenden Einzelindikatoren besteht, zu bewerten, gehört in den Bereich des „Edutainments“. Auch die extreme Informationsverdichtung – wie sie für

Rankings typisch ist – wird diesem Informationsbedarf immer weniger gerecht. Zukünftig wird sich wohl eher eine Arbeitsteilung herausbilden: einfache Rankings als Schlagzeilen, die eigentlich nur den Zweck haben, Akademiker auf ein Presseorgan aufmerksam zu machen und auf der anderen Seite methodisch solide Detailanalysen, die auch von der Darstellungsebene sich eher an Verortung in mehrdimensionalen Räumen und Netzwerkvisualisierungen orientieren. Das Siegereckchen aus dem Sport ist für die Abbildung von Lehr- und Forschungsleistungen unter solchen Bedingungen ein Anachronismus.

Literaturhinweise

Bayer, Christian (2004): Hochschul-Ranking. Vorschlag eines ganzheitlichen Ranking-Verfahrens, Duncker u. Humblot, Berlin.

Buela-Casal, Gualberto; Olga Gutiérrez-Martínez; María Paz Bermudez-Sánchez; Oscar Vadillo-Muñoz (2007): Comparative study of international academic rankings of universities. *Scientometrics*. DOI 10.1007/s11192-007-1653-8.

Centrum für Hochschulentwicklung – CHE (2006): CHE ForschungsRanking: <http://www.dasforschungsranking.de> (05.04.2007).

Deutsche Forschungsgemeinschaft – DFG (2006): Förder-Ranking 2006, Bonn.

Federkeil, Gero (2002): Some aspects of ranking methodology. The CHE-Ranking of German Universities, *Higher Education in Europe*, Vol. 27, No. 4, S. 389 – 397.

Federkeil, Gero (2006): Hochschulrankings – Orientierung für Studierende und Vergleichsmaßstäbe für Hochschulen. Schweizerische Gesellschaft für Verwaltungswissenschaften SGVW.

http://www.sgvw.ch/sektor/news/archiv/d/060926_che_ranking_federkeil.php (05.04.2007).

Gauffriau, Marianne, Peder Olesen Larsen (2005): Counting methods are decisive for rankings based on publication and citation studies, in: *Scientometrics*, Vol. 64, No. 1, S. 85 – 93.

Hornbostel, Stefan (1997): Wissenschaftsindikatoren. Bewertungen in der Wissenschaft, Opladen, Westdeutscher Verlag.

Hornbostel, Stefan (2001): Hochschulranking: Beliebigkeit oder konsistente Beurteilungen? Rankings, Expertengruppen und Indikatoren im Vergleich, in: Müller-Böling, Detlef, Stefan Hornbostel, Sonja Berghoff (Hrsg.): Hochschulranking. Aussagefähigkeit, Methoden, Probleme. Tagungsdokumentation, Gütersloh, Verlag Bertelsmann Stiftung, S. 7 – 42.

Hornbostel, Stefan (2006): Leistungsmessung in der Forschung. Von der Qualitätssicherung der Lehre zur Qualitätsentwicklung als Prinzip der Hochschulsteuerung. Beiträge zur Hochschulpolitik 1, S. 219 – 228.

Institute for Higher Education Policy – IHEP (Hrsg., 2007): College and University Ranking Systems: Global Perspectives and American Challenges. IHEP, Washington/DC.

Jobbins, D. (2002): The Times/The Times Higher Education Supplement. League Tables in Britain: An insider's view, Higher Education in Europe, Vol. 27, No. 4, S. 383 – 388.

Kirp, David L. (2003): Shakespeare, Einstein and the Bottum Line. The Marketing of Higher Education. Cambridge/Mass.

Kromrey, Helmut (2001): Informationsgehalt und Validität von Studierendenbefragungen als Evaluation, in: Müller-Böling, Detlef, Stefan Hornbostel, Sonja Berghoff (Hrsg.): Hochschulranking. Aussagefähigkeit, Methoden, Probleme. Tagungsdokumentation, Gütersloh, Verlag Bertelsmann Stiftung, S. 43-64.

Lebherz, Carmen, Cordula Mohr, Matthias Henning, Peter Sedlmeier (2005): Wie brauchbar sind Hochschul-Rankings?, in: Zeitschrift für Pädagogik, 50. Beiheft, S. 188 – 208.

Ledesma, J. R. (2002): Latin America: a different reality, Higher Education in Europe, Vol. 27, No. 4, S. 467 – 474.

Lücke, Nicole (2006): Der Preis ist heiß. DUZ 08/2006, S. 14 – 15.

Merisotis, Jamie P. (2002): Summary Report of the Invitational Roundtable on Statistical Indicators for the Quality Assessment of Higher/Tertiary Education Institutions: Rankings and League Table Methodologies. Higher Education in Europe, Vol. 27, No. 4, S. 475 – 480.

Merisotis, J. P., Sadlak, J. (2005): Higher education rankings: Evolution, acceptance, and dialogue, Higher Education in Europe, 30, S. 97 – 101.

Neidhardt, Friedhelm (2006): Forschungsevaluation, in: Soziologie, Jg. 35, Heft 4, S. 419 – 425.

Newman, Frank, Lara Couturier und Jamie Scurry (2004): The Future of Higher Education. Rhetoric, reality and the Risks of the Market. San Francisco.

Nickl Roger (2006): Wer ist die Beste im ganzen Land?, in: unijournal 2/06, Rektorat der Universität Zürich.

Research Assessment Exercise (RAE): <http://www.rae.uk> (05.04.2007).

Stölting, Erhard (2003): Rankings – sportlich gesehen. Eine Analyse aus soziologischer Sicht, in: Forschung und Lehre 10/03, S. 131 – 133.

The Times Higher Education Supplement – THES (2006): World University Rankings <http://www.thes.co.uk/worldrankings> (05.04.2007).

Van Raan, Anthony F. J. (2004): Measuring Science, in: Moed, Henk, F., Wolfgang Glänzel und Ulrich Schmoch (Hrsg.): Handbook of Quantitative Science and Technology Research. Dordrecht, Kluwer.

Van Raan, Anthony F. J. (2005): Fatal attraction: Conceptual and methodological problems in the ranking of universities by bibliometric methods, in: Scientometrics, Vol. 62, No. 1, S. 133 – 143.

Weingart, Peter (2005): Impact of bibliometrics upon the science system: Inadvertent consequences?, in: Scientometrics, Vol. 62, No. 1, S. 117 – 131.